**Evaluate Techniques for Wifi Locationing**

MAIN OBJECTIVE

Using "wifi fingerprinting" to determine a person's location in indoor spaces.

EXPLORATORY ANALYSIS

DATA GIVEN

Large database of wifi fingerprints for a multi-building industrial campus with a location (building, floor, and location ID) associated with each fingerprint.

The main characteristics of the database are:

It covers a surface of 108703m2 including 3 buildings with 4 or 5 floors depending on the building.

The number **of different places (reference points) appearing in the database is 933.**

**21049 sampled points** have been captured: **19938** for training/learning and **1111** for validation/testing.

Dataset independence has been assured by taking Validation (or testing) samples 4 months after Training ones.

The number of **different wireless access points** (WAPs) appearing in the database is **520**.

Data were collected by more than **20 users using 25 different models** of mobile devices (some users used more than one model).

VARIABLES:

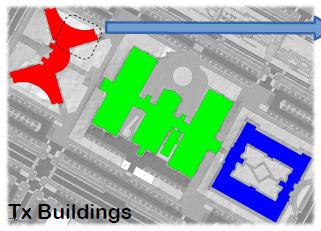
* **WAP 001-520:**

-100dBm is equivalent to a very weak signal, whereas 0dBM. The database includes 520 WAPs identified by the MAC address.

These addresses have been alphabetically sorted. the average number of WAPs scanned in each capture is 17:92.

The main factors that affect to the number of WAPs reported by a WiFi scan are location, the phone model (Android version and hardware) and how the device is held,

.

* **LONGITUDE /LATITUDE/ FLOOR**

Represented in each sample/capture by means of three values in each record

* **"BUILDING ID"**

[0-2] corresponds to the building in which the capture was taken

* **"SPACE ID"**

Integer value that is used to identify the particular space

* **"RELATIVEPOSITION"**

[0-1] Inside or outside

* **"USERID"/"PHONEID"**

USER ID is used to represent the 18 different users who participated in the procedure to generate the training samples.

PHONE IDcontains an integer value to represent the Android device used in each capture.

* **"TIMESTAMP"** time in Unix time format

VALIDATION DATASET

* 14 users installed theapplication on their Android devices and executed it during 20 minutes (approximately) in each of the three Tx buildings.
* sending the required information (only WAPs detected and RSSI levels) to a centralized server, and it gets a point inside a building (given by its longitude, latitude, floor) from the server

PREDICTION

BUILDING

Sample n=200

|  |  |  |
| --- | --- | --- |
|  |  |  |
| RANDOM FOREST |  |  |
|  |  |  |
|  |  |  |

RANDOM FOREST MTRY =6

Confusion Matrix and Statistics

predicted\_building 0 1 2

0 535 0 0

1 1 307 1

2 0 0 267

Overall Statistics

Accuracy : 0.9982

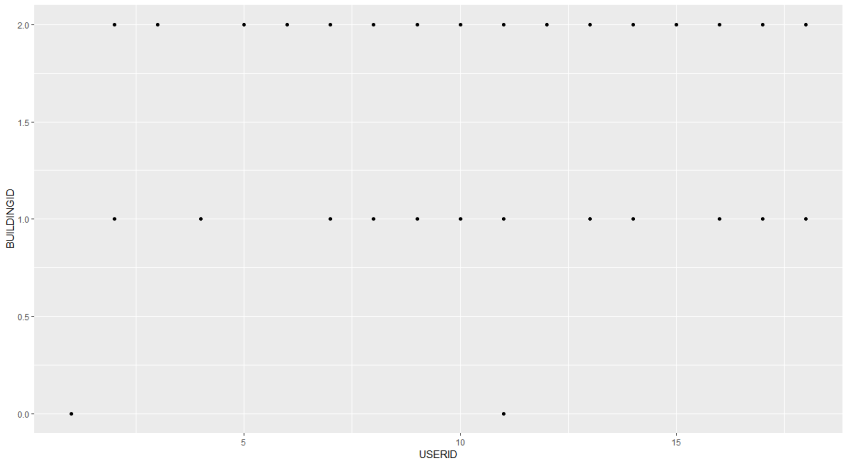
95% CI : (0.9935, 0.9998)

No Information Rate : 0.4824

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9972

DATA EXPLORATION

* Reduce the WAPS deleting all the ones not being used (the ones with a mean of = 100)
* All users have data
* Not all phones have been used (only 16)
* Phone 14 is being used by 3 users.
* Buildings:
* **1 Math and computer**-> 5000 samples and only 2 users
* **2 Technology & Science -> 5000 samples**
* **3 T&S research -> 9000 samples**